DOCUMENT RESUME

ED 445 025 TM 031 630

AUTHOR Conley, David T.
TITLE Who Is Proficient: The Relationship between Proficiency Scores and Grades.
PUB DATE 2000-04-24
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Academic Standards; College Entrance Examinations; *Evaluators; *Grades (Scholastic); Grading; High Schools; Higher Education; *Scoring; *Secondary School Teachers
IDENTIFIERS *External Evaluation; Oregon

ABSTRACT

This paper examines the relationship between the grades teachers give their students and the scores external raters give the same students' work when using a common set of standards and criteria. Data were collected from 78 high school teachers participating in the Proficiency-Based Admission Standards System (PASS) project to establish standards for the Oregon University system. These teachers collected work from approximately 2,200 students over 1 academic year. Each teacher focused on one or two of the defined "proficiencies" in English or mathematics, and each collected multiple pieces of work from each of their students in targeted classes. Each teacher brought five to nine of these collections to be reviewed by trained reviewers, either other teachers or university professors. Findings suggest that proficiency scores are measuring something related to, but not the same as, grades. The degree of correlation between the two does suggest a relatively linear relationship between scores and grades, and that the relationship is in the expected direction. Statistical analyses also indicated that the distributions of proficiency scores and grades were statistically different. Stepwise regression analyses found little relationship between the grading system a teacher used and whether or not a student was proficient. These findings make a case for the existence of grade inflation. (Contains 20 tables and 10 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 445 025

TM

Who is Proficient: The Relationship between Proficiency Scores and Grades

David T. Conley, Ph.D.[1]
Associate Professor, College of Education
University of Oregon
Eugene, OR 97403
541-346-2445
conley@oregon.uoregon.edu
AERA, New Orleans — April 24, 2000

This study examines the relationship between the grades teachers give their students and the scores external raters give the same students' work when using a common set of standards and criteria. The study compares teachers' idiosyncratic grading systems with judgments derived from a standards-based scoring system. Stated differently, what is the relationship between student grades on an A-F scale and student proficiency scores on a 5-point scale? To further understand this relationship, teacher grading systems are considered in relation to student proficiency scores.

The study is primarily exploratory in nature. It seeks to determine if proficiency judgments are different from grades awarded. It also considers if proficiency scores differ by student grade level or subject area (English and mathematics). These are important questions because many schools and a number of states are implementing scoring systems where teachers are trained to judge student work by applying standards and criteria to reach judgments within a common scoring framework. These systems require considerable training to enable teachers to attain adequate reliability levels. Such systems require maintenance and support to ensure the validity of standards and criteria and the continuing reliability of teacher judgments. The relationship between these new assessment technologies and the more familiar, institutionalized system of grades needs to be understood better. Raising these issues helps advance current discussions about and understanding of assessment policy and practice and the relative utility of various assessment methods.

This study takes place within the state of Oregon, which has adopted proficiency-based university admission standards for students admitted beginning fall, 2005 to the state's public universities. These standards are being piloted at 50 high schools, which are charged primarily with field testing the assessment methods needed to make proficiency-based admission decisions. This study generates baseline data on the relationship between proficiency scores and grades. Students in this study who choose to attend Oregon public universities will be followed as they progress in the university to determine further the relationship between proficiency scores, high school grades, and subsequent university performance. This study will help determine what role grades should

[1] Dr. Conley is also Executive Director of the Proficiency-based Admission Standards System (PASS), Oregon University System

TM031630

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

D T. Conley

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

play in the admissions process once a transition to proficiency-based admission is completed. At a more basic level, this study helps determine if a proficiency-based assessment system measures the same or different constructs as grade, or if it measures the same constructs with greater or lesser precision.

Perspectives/theoretical referents

This is not a simple study of concurrent validity or scorer reliability, although those issues are considered. It is focused more precisely on the role of teacher judgment within two separate referent system. In that sense, it examines the role of teacher judgment as a component of the classic reliability-validity formula, rather than considering it separately, as for example a variable in a reliability coefficient.

Marzano (1994) has considered whether teachers can make judgments on some proficiencies without administering performance assessments. He has found some proficiencies easier than others to judge. Marzano notes that while performance assessments have high face validity, they do not necessarily measure what they purport to measure. Simply developing common tasks, standards, and scoring methods is no guarantee of adequate validity or reliability.

Caroline Gipps (1994) raises important issues about reliability in a standards-based system that employs criterion-referenced assessment. She notes that in a criterion-referenced system, the concept of reliability in its traditional sense is not appropriate, since such measures are based on correlation techniques that assume high levels of discrimination between pupils and a wide range of scores. Since the goal of criterion or standards-referenced assessments is not to generate a normal distribution of scores, the range of scores will generally be narrow and bunched near the desired performance level. As a result, alternative approaches need to be employed to evaluate the consistency of measurement and the stability of the classification system itself.

Additionally, the limitations of grading have been noted with increasing frequency. Gusky (1994) affirms the importance of relating grading to learning criteria, while Ornstein (1994) observes that the more detailed the reporting method and the more analytic the process, the more likely subjectivity will influence results. Teachers struggle to learn how to integrate new performance assessments with traditional grading systems (Seeley, 1994). Grade inflation has been noted at all levels of the educational system, from high school to universities (Gose, 1997; Ziomek & Svec, 1995). Meanwhile, many schools move beyond grading with little assurance their new methods are an improvement on the old ones (Minneapolis Star Tribune, 1998).

Methodology

Data were collected in the spring of 1999 from 78 of the 100 teachers from 50 high schools participating in the Proficiency-based Admission Standards

System (PASS) project to establish new standards for the Oregon University System. These teachers collected student work from approximately 2,200 students over one academic year. Each teacher focused upon one or two "proficiencies" in either English or math. These proficiencies were statements of the knowledge and skills students were expected to possess within English and math in order to be ready for university admission. Each proficiency had between two and seven criteria that were used in combination with the proficiency itself to judge the adequacy of the student work collected. Table 1 contains a statement of the proficiencies and scoring criteria for each proficiency in English and math.

Table 1: English and math proficiencies

English Proficiencies and Criteria

A: Read from a Variety of Literary Genres and Periods: Read and respond to a broad selection of literature from a variety of historical periods, cultures, literary perspectives, and genres, including poetry, novels, short stories, essays, and drama; understand the characteristics of literary genres, periods, and movements.

- A1: Breadth and Depth of Literary Experience: Read and respond to works of recognized literary merit from a variety of historical periods, cultures, and genres.

B: Interpret Literary Works: Analyze literary forms, elements, devices, and themes to interpret and critique literary texts, performances, and media.

- B1: Analysis of Literary Elements and Devices Recognize, examine, and understand the uses and effects of literary elements, rhetorical devices, and themes within and among literary works.
- B2: Interpretation and Use of Textual Evidence: Use textual evidence to develop and support an interpretation of a literary work.
- B3: Criticism: Use introductory ideas and approaches of literary criticism in analyzing and critiquing a literary work.

C: Analyze Relationships of the Humanities & Human/Social Experience Explain how literature and the humanities reflect, influence, and comment upon human experiences and societal assumptions, traditions, structures, and changes.

- C1: Understanding of Contextual and Biographical Influences: Explain how works from the humanities are influenced by historical, social, cultural, political, literary, or creative contexts and individual experiences.
- C2: Understanding of Social/Cultural Representations: Examine how works from the humanities characterize, individuals, groups, and cultures.
- C3: Understanding of Social/Cultural Commentary: Explain social/cultural perspectives, themes, and commentary, and examine techniques used to promote or critique social change in works from the humanities.

D: Conduct Inquiry and Research: Conduct inquiry and research, using a variety of primary and secondary sources and informational resources to investigate questions and topics, gather and synthesize information, and create and communicate knowledge.

- D1: Research Process: Identify and frame topics, questions, and purposes for inquiry; plan and conduct research.
- D2: Analysis of Information Sources: Locate and interpret varied information sources; distinguish among facts, supported inferences, and opinions; evaluate information.
- D3: Use of Researched Information: Use, integrate, and cite researched information and evidence.

E: Communicate in Oral, Visual, and Written Forms: Use oral, visual, written, and multi-media communication forms to convey information and ideas for a variety of purposes, audiences, and contexts.

- E1: Use of Oral, Visual, and Written Forms: Use and integrate oral, visual, written, or multimedia forms to communicate ideas in ways appropriate to topic, context, audience, and purpose.
- E2: Organization of Presentations: Organize oral, visual, or multimedia presentations in clear, coherent sequences appropriate to topic, context, audience, and purpose.
- E3: Use of Language and Techniques: Use the languages, techniques, and conventions of various communication forms to communicate ideas.
- E4: Analysis of Oral, Visual, Written, and Multimedia Communications: Analyze and evaluate oral, visual, and written/media communications, considering topic, context, audience, purpose, delivery, and language.

F: Write for Varied Purposes: Write to discover and convey meaning, using effective processes to produce writing which is thoughtful, fluent, organized, coherent, and clear.

- F1: Quality of Thinking (Ideas and Content): Develop support, and convey clear, focused, and substantive ideas in ways appropriate to topic, context, audience, and purpose.
- F2: Organization and Coherence (Organization): Organize writing in clear, coherent sequences, making connections and transitions among ideas, paragraphs, and sentences.
- F3: Style and Technique (Sentence Fluency and Word Choice): Use and vary sentence structures, word choices, and writing voice to achieve clear and fluent writing.
- F4: Conventions and Format (Conventions and Citing Sources): Use correct spelling, grammar, punctuation, capitalization, paragraph structure, sentence construction, formatting, and, when appropriate, citations.

F5: Purposes, Modes, and Forms
Write for varied purposes in a variety of modes and forms.

F6: Writing Process
Use effective processes to generate, compose, organize, revise, and present writing.

Math Proficiencies and Criteria

A: Perform Algebraic Operations: Use algebraic operations and mathematical expressions to solve equations and inequalities including, but not limited to, exponentials and logarithms.

A1: Solving Equations and Inequalities: Solve equations and inequalities numerically, graphically, and/or algebraically.

A2: Use of Matrices: Use matrices to organize information and to solve systems of equations.

B: Use Functions to Understand Mathematical Relationships: Use patterns and functions to represent relationships between variables and to solve problems; interpret and understand the connections among symbolic, graphic, and tabular representations of functions. (Note: Students should demonstrate proficient understanding of linear, quadratic, general polynomial, inverse variation, and exponential functions, and familiarity with logarithmic and trigonometric functions.)

B1: Representation and Recognition of Functions: Represent functions using and translating among words, tables, graphs, and symbols; recognize and distinguish a variety of classes of functions.

B2: Analysis of Functions: Understand and analyze features of a function and limitations on the domain of a function.

B3: Use of Functions as Models: Model situations and solve problems using a variety of functions.

C: Use Geometric Concepts and Models: Represent and solve problems with two- and three-dimensional geometric models, properties of figures, analytic geometry, and trigonometry.

C1: Use of Coordinate Geometry
Represent, interpret, and analyze geometric figures and properties using drawings, models, and/or the Cartesian coordinate system.

C2: Use of Plane Geometry: Use properties and relationships of geometric figures to analyze and model natural and constructed forms.

C3: Direct and Indirect Measurement: Use geometry and trigonometry to determine measurements.

C4: Use of Geometric Models: Use geometric relationships, spatial reasoning, & models to solve problems.

D: Use Probability and Statistics to Collect and Study Data: Use probability and statistics in the study of various disciplines, situations, and problems;

understand and apply valid statistical methods and measures of central tendency, variability, and correlation in the collection, organization, analysis and interpretation of data.

D1: Use of Probability Models: Use experimental or theoretical probability to represent and interpret situations or problems involving uncertainty.

D2: Statistical Investigation: Design and conduct statistical experiments, simulations, or surveys; collect data.

D3: Organization and Use of Data: Create, interpret, and analyze charts, tables, and graphs to display data, draw inferences, make predictions, and solve problems.

D4: Interpretation of Data : Analyze data using descriptive and inferential statistics; interpret statistical results.

E: Estimate and Compute: Use computation, estimation, and mathematical properties to solve problems; use estimation to check the reasonableness of results, including those obtained by technology.

E1: Estimation
Estimate solutions and determine if the results are accurate and reasonable.

E2: Computation: Perform numeric and algebraic calculations on real numbers, expressions, and matrices, using appropriate methods and tools, including technology.

E3: Verifying Results: Use estimation to verify results and identify potential errors when using technology.

F: Solve Mathematical Problems: Apply mathematical problem-solving strategies to problems from within and outside mathematics; devise, implement, and evaluate processes and solutions; select and use appropriate models, operations, and technologies.

F1: Formulating and Understanding: Understand and formulate problems; select or provide relevant information; use mathematical concepts, models and representations.

F2: Processes and Strategies: Consider and choose among various strategies, algorithms, models, and concepts to devise and carry out solutions.

F3: Communication: Represent and communicate processes, solutions, ideas, and conclusions; use correct mathematical terminology, symbols, and notation.

F4: Verification: Evaluate processes, strategies, calculations, and solutions to verify reasonableness; explore alternative approaches, extensions, and generalizations.

G: Reason Mathematically: Formulate and test mathematical conjectures (i.e., make generalizations from observations); draw logical conclusions from given or known information; follow and judge the validity of mathematical arguments and proofs.

G1: Mathematical Reasoning: Formulate and test mathematical conjectures and conclusions.

G2: Mathematical Arguments: Follow, evaluate, and develop mathematical arguments and proofs.

Teachers collected multiple pieces of work from each of their students in the targeted classes. These pieces of work were assembled into "collections of evidence" designed to demonstrate proficiency in the targeted area. Teachers were trained to judge student work via a three-part process; first, they developed assessment plans in October of the preceding year, which were reviewed to ensure quality; second, three months later, they brought samples of student work to a scoring session for cross-scoring; and, third, they brought complete collections to a scoring session for cross-scoring by other teachers in May.

Teachers were instructed to bring between five and nine collections from among all of their students to a scoring session held on May 21, 1999. Oregon University System staff, not the teacher, selected the students at random from class rosters submitted by the teachers.

Each student's collection of work was placed in a folder and given a coded number. All information that might identify the student, teacher, or school was removed from the work. Each collection was then reviewed anonymously by a minimum of three trained reviewers who could be either teachers or university professors. Each reviewer followed a common process for scoring collections of student work and utilized a common form for reaching a judgment about the collection on a five-point scale (1-5) Teachers did not score their own students' work.

Table 2 contains an excerpt from the scoring sheet each scorer used. Table 3 describes the five-point scale. Note that a score of 3 on each proficiency is the level needed to meet university entrance standards.

Table 2: Process for scoring student work

STEP 1: Determine Sufficiency of Evidence and Proficiency of Performance	
SUFFICIENCY: To determine the sufficiency of evidence, answer the following questions:	PROFICIENCY: To determine proficiency of performance, apply the following decision rules:
Does the collection sufficiently represent the standard? The collection addresses the range of criteria or allows inferences about criteria not addressed.	Exceeds the Standard The collection is above the description of proficient performance and allows inferences about knowledge and skills.
Have there been sufficiently varied opportunities and conditions for assessment? The collection represents ample assessment variety for	Meets the Standard The collection is consistent with the descriptions of proficient performance and allows

demonstrating proficiency.	inferences about knowledge and skills.
Is there sufficient evidence to be confident that the work represents the student? Indicates that the work is the student's own performance.	Does Not Meet the Standard The collection does not indicate performance as described at the proficient level.

STEP 2: Assign a Summary Judgment Score Determine Sufficiency of Evidence and Proficiency of Performance Note: Sufficiency and proficiency are interrelated. Consider both before making both judgments.
If there is sufficient evidence to make a confident judgment AND if the student's work consistently meets and regularly exceeds the criteria, then the summary judgment score is 5 or 4.
If there is sufficient evidence to make a confident judgment AND if the student's work meets the criteria, then the summary judgment score is 3.
If there is insufficient evidence to make a confident judgment OR if the student's work does not meet the criteria, then the summary judgment score is 2 or 1:

Table 3: Five-point scale for summary judgments of student work collections

Performance Characteristics of Performance/Decision Rules	
(E) Exemplary*	The collection demonstrates an exemplary mastery of the proficiency and exhibits exceptional intellectual maturity or unique thinking, methods, or talents.
(H) High-level Mastery of the Proficiency*	The collection demonstrates mastery of the proficiency at a level higher than entry-level college coursework.
(M) Meets the Proficiency	The collection demonstrates the student is prepared for entry-level college coursework
(W)Working Toward the Proficiency	The collection approaches readiness for entry-level college coursework. The level of performance may be improved by: • providing a broader variety of opportunities and conditions of assessment; • providing sufficient evidence to address the range of criteria for the proficiency; • enrolling in more classes that target this proficiency.
(N) Not Meeting the Proficiency	The collection contains evidence that the student is not prepared to do entry-level college coursework.

The second element of the study was the grades students received in the course in which they prepared their collection of work. This information was reported by the teachers in the form of letter grades on a traditional A-F scale. Teachers were asked to identify the means by which they arrived at grades in the course. They were asked to apportion 100 points among a variety of grading options based on how important each was in determining the students' grades. Table 4 contains the options presented.

Table 4: Teacher Grading Method

Grading method:	% Importance
Tests	
Final	
Homework	
Research paper or term paper	
Participation	
Attendance	
Individual project(s)	
Group project(s)	
Assignments completed in class	
Other:	
Other:	
Total:	*100%*

Teachers were also asked to report the proficiency or proficiencies they targeted in the course, and, if more than one proficiency was targeted, the relative emphasis placed on each proficiency.

Findings

Table 5 summarizes the number of collections of work that were submitted by student grade level in high school Teachers were asked to target juniors and seniors, but were not prohibited from including freshmen and sophomores. Students tend to be placed into English classes on the basis of year in school. However, in mathematics, students often accelerate during middle school, resulting in more freshmen and sophomores in college-preparation mathematics courses. This may explain the slightly higher average number of work collections submitted by freshmen and sophomores in mathematics.

Table 5: Average number of work collections by student grade level

Avg. # of collections per grade level:	English	Mathematics
Freshmen	2.5	6.5
Sophomores	8.4	11.8
Juniors	19.9	14.1
Seniors	12.5	12.0

Table 6 illustrates the distribution of proficiency scores, which reflects a more normalized distribution. The distribution of grades as shown in Table 7 indicates a higher concentration of A's, followed by B's, then C's. This is consistent with grading practices in college-bound classes. Although this pattern did not hold over every class, it was common to most courses.

Table 6: Distribution of summary judgment scores on 1-5 scale

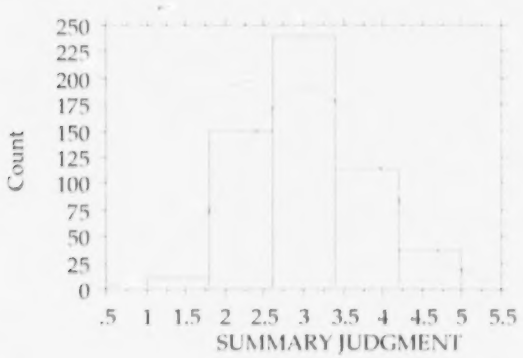
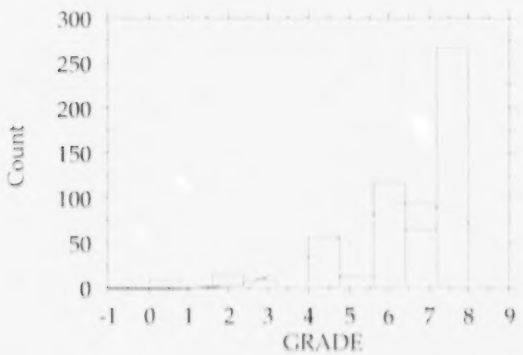


Table 7: Distribution of grades on 1-8 scale



Scale: 8=A, 7=A-/B+, 6=B, 5=B-/C+, 4=C, 3=C-/D+, 2=D, 1=D-

Table 8 shows the relationship between the grades students received in the mathematics class in which they produced their collection of work that was subsequently judged and the proficiency area in which the collection was produced. Most students produced collections in Math Proficiency F (Solve Mathematical Problems), in large measure because problem solving is an area upon which students are tested via a performance task in grades 3, 5, 8, and 10. Teachers were more familiar with problem solving as a task for which student work could be produced. The next most-frequently targeted proficiency was Math Proficiency B (Use Functions to Understand Mathematical Relationships), followed by Math Proficiency C (Use Geometric Concepts and Models). These two proficiencies align themselves well with existing courses, particularly Algebra and Geometry.

Table 8: Grade received by proficiency area: Math

	D	D+/C-	C	C+/B-	B	B+/A-	A	Totals
Math A	1	0	1	0	4	4	2	12
Math B	1	5	10	1	11	9	23	60
Math C	2	0	1	4	2	6	25	40
Math D	0	0	1	1	2	4	4	12
Math F	3	4	11	2	19	10	51	100
Totals	7	9	24	8	38	33	105	224

Table 9 provides the same information as Table 8, but for English, where teachers concentrated on English Proficiencies B (Interpret Literary Works) and F (Write for Varied Purposes) primarily. As in mathematics, these proficiencies were easier to address without making major changes in curriculum and instructional methods. Grades here were also negatively skewed, with the largest concentration in A's, followed by B's and C's. This group of students also represented college-bound students who were achieving A's at a high rate.

Table 9: Grade received by proficiency area: English

	D-	D	D+/C-	C	C+/B-	B	B+/A-	A	Totals
Eng. A	1	2	0	5	0	6	2	12	28
Eng. B	0	4	1	9	2	32	12	53	113
Eng. C	0	0	0	0	0	4	1	7	12
Eng. D	0	0	0	1	1	7	4	18	31
Eng. E	0	0	0	1	0	0	1	3	5
Eng. F	0	2	0	16	1	30	12	68	129
Totals	1	8	1	32	4	79	32	161	318

Table 10 provides a more detailed breakdown of the grades students received by mathematics course. Proportions of students earning each grade category is similar in each course, with the exception of A.P. Calculus and Math Analysis, each of which had very small n's. Algebra 2 also had a wider spread of grades than most other courses.

Table 10 Grades by math class

	D	D+/ C-	C	C+/B-	B	B+/ A-	A	Totals
Adv. Geometry	0	0	0	0	0	0	12	12
Advanced Algebra	0	0	1	0	4	1	6	12
Algebra 1	0	0	0	0	2	0	4	6
Algebra 2	3	0	4	0	7	0	10	24
Algebra/Geom. 2	0	0	1	0	2	0	2	5
AP Calculus	0	4	4	0	0	0	2	10
Functions & Trig	0	0	0	0	0	5	6	11
Geometry	2	0	2	5	3	7	18	37
IB Calculus	0	0	2	0	4	2	4	12
Integrated Math 3	0	1	1	0	1	1	8	12
Interactive Math 2	0	0	0	1	0	2	1	4
Math Analysis	0	0	0	0	1	4	1	6
Math Analysis 4	2	2	0	2	0	4	2	12
Pre-calculus	0	2	6	0	13	7	28	56
Trigonometry	0	0	3	0	1	0	1	5
Totals	7	9	24	8	38	33	105	224

Course titles in English are so varied that it is difficult to make any generalizations about grades by specific course. Table 11 illustrates the range of course titles encountered.

Table 11 Grades by English class

	D-	D	D+/ C-	C	C+ /B-	B	B+/ A-	A	Totals
Adv Junior English	0	0	0	0	0	3	0	3	6
Adv. Composition	0	0	0	1	0	2	0	3	6
Adv. English 10	0	0	0	0	0	1	0	5	6
American Lit	0	1	0	1	0	5	2	14	23
American Literature	0	0	0	0	0	0	0	5	5
American Studies	0	0	0	0	1	0	4	1	6
American Writers	0	0	0	0	0	1	0	4	5
AP English	0	1	0	0	0	5	6	3	15
AP English Lit	0	0	0	1	0	3	0	2	6
AP Literature	0	0	0	0	0	10	0	14	24
College Comp.	0	0	0	0	0	3	0	3	6
College Writing	0	0	0	0	0	0	0	2	2
Comp-Literature	0	0	0	3	0	0	3	0	6
CP English 11	0	0	0	0	0	1	6	9	16
CP English 12	0	0	0	2	0	1	0	3	6
English 7-8	0	2	0	2	0	2	4	2	12
English	0	2	0	6	0	12	0	14	34
English 11	0	0	0	1	0	2	1	1	5
English 12	1	0	1	2	0	3	0	1	8
English 3	0	0	0	1	0	0	1	3	5

English 9	0	0	0	1	0	1	0	6	8
English I	0	0	0	0	0	1	2	2	5
Honors Amer. Exp	0	0	0	0	0	1	0	2	3
Honors English	0	0	0	1	0	3	0	6	10
Honors Jr. English	0	0	0	2	0	0	0	10	12
Honors Lit.	0	0	0	0	0	2	1	3	6
Humanities	0	0	0	0	0	0	0	6	6
Independent Study	0	0	0	0	0	0	0	2	2
JR IB English	0	0	0	0	0	4	0	2	6
Junior English	0	0	0	1	0	2	0	2	5
Language Arts	0	0	0	2	0	6	0	4	12
Power English	0	0	0	0	0	1	0	3	4
Research Writing	0	0	0	4	2	0	2	4	12
Senior English	0	2	0	1	0	2	0	1	6
Sophomore English	0	0	0	0	0	0	0	6	6
Women's Lit	0	0	0	0	1	0	0	6	7
World Literature	0	0	0	0	0	2	0	4	6
Totals	1	8	1	32	4	79	32	161	318

Table 12 explores the relationship between grades and proficiency scores. Correlations were calculated using both Person Product-Moment and Spearman Rank Order. Results were very similar using both methods.

The key observations are that the correlation between English proficiency score and grade (.474) and between math proficiency score and grade (.452) were very similar. While correlations varied by individual courses, few true outliers exist.

Math Proficiency B (Use Functions to Understand Mathematical Relationships) correlated most highly (.647) of any proficiency with more than 25 cases. Calculus and Math Analysis were the courses that correlated most highly, while all other courses except Geometry demonstrated strong correlations.

Table 12: Relationship between grades and proficiency scores

Proficiency/Area	Correlation between Summary Judgment and Grade in Class	Average proficiency score	Average grade	# of Cases
Overall	.459	3.03	6.60 (B+)	550
Overall-Math	.452	3.05	6.46 (B)	234
Math Proficiency A	.691	3.00	6.17 (B)	12
Math Proficiency B	.647	3.23	6.05 (B)	62
Math Proficiency C	.347	2.83	7.08 (A-)	40
Math Proficiency D	.575	3.50	6.75 (B+)	12
Math Proficiency F	.374	2.95	6.51 (B+)	102
Algebra	.413	2.85	6.31 (B)	48
Calculus	.750	3.42	5.08 (B-)	24

Geometry	.196	2.82	6.86 (B+)	55
Math Analysis	.822	2.94	5.89 (B-)	18
Pre-calculus	.632	3.18	6.80 (B+)	56
Trigonometry	.401	3.29	6.41 (B)	17
Overall-English	.474	3.02	6.71 (B+)	316
English Proficiency A	.729	3.21	5.90 (B-)	29
English Proficiency B	.393	2.81	6.67 (B+)	113
English Proficiency C	.816	3.67	7.25 (A-)	12
English Proficiency D	.324	3.27	7.20 (A-)	31
English Proficiency F	.499	3.05	6.73 (B+)	131
Junior English	.393	2.98	7.00 (A-/B+)	51
Literature	.466	3.15	7.05 (A-)	101
Writing	.576	2.88	6.34 (B)	32
Senior English	.585	2.67	4.95(C+)	21
Junior English	.393	2.98	7.00 (A-/B+)	51
Sophomore English	.512	3.17	7.83 (A)	12
AP English	.709	3.44	6.84 (B+)	45
Honors English	.574	2.94	7.20 (A-)	31

Table 13 presents the average proficiency score students received for their collection of work grouped by the grade they received in the class in which they were enrolled when they completed their collection of work. The relationship is linear with each lower grade receiving a lower average proficiency score with the exception of the 12 students who received B- or C+ grades. Their average score was only slightly higher than those receiving a grade of B.

Table 13: Average proficiency score by letter grade received in class

Grade in class	Proficiency score	Standard deviation	Number of cases
A	3.44	.910	266
A-/B+	3.05	.837	65
B	2.66	.632	117
B-/C+	2.75	.452	12
C	2.43	.599	56
C-/D+	2.20	.789	10
D	2.13	.516	15
D-	1.00	0.000	1

Table 14 considers proficiency scores by student grade level in high school. Average proficiency scores drop between 12th graders and 11th graders, are marginally higher for 10th graders, then drop markedly for 9th graders.

Table 14: Average proficiency score by grade level

Grade in class	Proficiency score	Standard deviation	Number of cases
12th grade	3.14	.901	209
11th grade	3.00	.967	188
10th grade	3.09	.770	71
9th grade	2.77	.865	82

Table 15 catalogs the importance that teachers placed on certain methods when arriving at the grade for students by subject area. Teachers were asked to determine what percentage of the grade was determined by each of nine grading methods. The most marked differences between math and English grading systems were in the importance of tests and the final, which were more important in math, and of research or term papers, which were more important in English.

Table 15: Elements of teacher grading systems

Grading method:	Percent Importance-English	Percent Importance-Math	Combined Percent Importance
Tests	20.50	45.76	31.52
Assignments completed in class	17.52	4.39	11.83
Homework	17.39	19.47	18.30
Research paper or term paper	11.67	1.28	7.09
Individual project(s)	10.46	6.42	8.70
Other:	8.64	4.32	6.76
Final	6.01	12.66	8.91
Participation	3.90	3.19	3.58
Group project(s)	3.23	1.62	2.53
Attendance	.74	.35	.57
Other:	.23	.53	.36

Table 16 investigates whether the group of students judged proficient is different from the group judged not proficient. The results confirm that the populations are indeed different, that students who are judged proficient are not necessarily the same students with high grades. This nonparametric test reinforces the notion that although there is a relationship between grades and proficiency scores, as suggested by previous data, the two groups are different in a statistically significant fashion. Students whose work collections were judged proficient (score of 3 or greater) received different grades from those whose collections were judged not proficient (score of less than 3) to a statistically significant degree.

Table 16: Kolmogorov-Smirnov Test for the relationship between grade received and whether work was judged proficient or not proficient

DF	2
Count, Not proficient	160
Count, Proficient	390
Maximum Difference	.382
Chi Square	66.074
P-Value	<.0001

Table 17 and Table 18 take the opposite look at the same data. The Kruskal-Wallis Test demonstrates that there is a difference between the proficiency score received and grade received. Students with high grades received proficiency scores that were different from those receiving low grades.

Table 17 Kruskal-Wallis Test for SUMMARY JUDGMENT

Grouping Variable: LETTER GRADE

DF	7
# Groups	8
# Ties	5
H	115.688
P-Value	<.0001
H corrected for ties	130.221
Tied P-Value	<.0001

8 cases were omitted due to missing values.

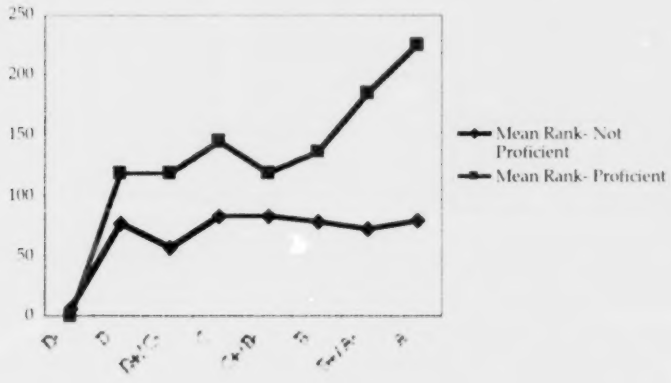
Table 18: Kruskal-Wallis Rank Info

	Count	Sum Ranks	Mean Rank
D-	1	6.000	6.000
D	15	1738.000	115.867
D+/C-	10	1436.000	143.600
C	56	9164.500	163.652
C+/B-	12	2706.000	225.500
B	117	24578.500	210.073
B+/A-	65	18051.000	277.708
A	266	89473.000	336.365

8 cases were omitted due to missing values.

Table 19 compares the mean ranks of students receiving proficient ratings with those receiving not proficient ratings. For the proficient group, mean rank and grade received are well related. Among students whose work was rated not proficient, the relationship is less linear. This demonstrates the differences in the two populations.

Table 19: Proficient/Not Proficient by Mean Rank from Kruskal-Wallis Test



The results from the multiple regression displayed in Table 20 demonstrate the lack of relationship between teacher grading system and student proficiency score received.

Stepwise regression revealed a relationships between homework and not proficient in English and homework and participation in math. These were the only variables that met the F-to-Remove criteria of the stepwise regression, demonstrating the weak relationship between grading system and proficiency score received.

Table 20: Multiple regression: Proficiency score and elements of teacher grading system

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	3.197	.248	3.197	12.879	<.0001
Tests	-.001	.003	-.023	-.321	.7483
Final	-.001	.004	-.020	-.373	.7090
Homework	-.010	.004	-.172	-2.767	.0059
Paper	-.003	.004	-.047	-.809	.4190
Participation	.018	.009	.107	2.069	.0390
Attendance	-.038	.020	-.093	-1.927	.0545
Ind. Projects	.002	.004	.027	.505	.6137
Grp Projects	.019	.008	.112	2.367	.0183
Class Assign	-.003	.003	-.052	-.822	.4113

Discussion

The comparison of student scores on collections of work with grades received in class result in correlations in the .45 range. This suggests that proficiency scores are measuring something related to but not the same as grades. This conclusion is reinforced by the results from the Kolmogorov-Smirnov Test (Table 16), and the Kruskal-Wallis Test (Table 17), each of which indicates that the distribution of proficiency scores and of grades are statistically different.

The stepwise regression analysis examines teacher grading systems and student proficiency scores and found very little relationship between the grading system a teacher used and whether or not a student was proficient.

What are we to make of these findings? Should we expect more correlation among these elements; grades, grading system, and proficiency score; or is it logical to expect only modest relationships between grades and proficiencies, and little or no relationship between grading system and proficiency score? Why are these measures not more highly inter-correlated?

At the least, the findings suggest that grades and proficiencies are in fact measuring different things to a significant extent. Lower correlations would have suggested separate constructs were being measured. Higher correlations might have hinted that proficiency scores duplicated what grades have come to measure. The middling correlations suggest a relationship, but not a duplication, between the two measures.

One possible explanation is that grades are reflections of a wider range of attributes than what is judged in the collections. The collection, by its design, is focused primarily on written assignments, unit tests, and individual projects. These are important elements of the grading process, but are not the sole elements. Homework in math and in-class assignments in English, in particular, make up a substantial portion of the grade, but neither type of work lends itself well to inclusion in a collection. And even though teachers said they gave little emphasis to attendance (less than one percent) and there was no category to indicate student behavior was considered in awarding grades, it can be reasonably assumed that these two dimensions did have an influence on grades. And "extra credit" was not taken into account, since, technically, it is not a formal dimension of teacher grading systems, but is frequently used in practice.

By contrast, the means by which collections are scored (anonymously) eliminates consideration of individual attributes, such as effort, special circumstances, race and ethnicity, or the halo effect, where a student who does well on one piece of work or in one area of the curriculum then benefits from higher marks on all pieces of work or in all areas of the curriculum.

The relationship between average score and letter grade (Table 13) demonstrates that there is a relatively linear relationship between score and grade, and that the relationship is in the expected direction. In other words, as grade increases, so does proficiency score. It is also worth noting that the average

score for students with a grade of A (3.44) is above the proficiency score of 3 required for admission (as is the score for the A-/B+ students, 3.05), but the average for students with a B (2.66) is not above the minimum proficiency level expected for admission. This suggests that, at least under the current scoring methods, students receiving a B are performing below the level deemed appropriate for success in entry-level university classes. However, some of these students are freshmen and sophomores, students who would not be expected to have reached that level. Table 14 provides some evidence for higher proficiency scores at higher grade levels, but additional analyses suggested the differences between proficient and not proficient by grade level was more modest than the differences in means suggest.

At least part of this phenomenon could be explained by the variation in grading systems. No two teachers had the same grading system, varying their weighting of the same nine elements. As a natural result, students in two different classes with the same title would have had to do well on different sorts of work to receive high grades. By contrast, the proficiencies required the same sorts of things of all students. Given that at this point grades are still more important than proficiency scores, students would have naturally placed more emphasis on activities that yielded a high grade. Particularly in those classes that emphasized homework as an important component of the grading system, it would have been possible to get a good grade while not necessarily developing or demonstrating many of the skills required by the proficiencies, which were oriented toward tests, papers, and other "demonstrations" of knowledge.

Part of the difference comes from the fact that the proficiency scoring system is focused on the difference between a 2 and a 3, which is, in essence, the difference between admission and rejection, and paid relatively less attention to distinctions between a level 3 and levels 4 and 5. This is evident in part by the much larger concentration of scores at the 3 level compared to grades, which demonstrated much larger concentrations at the A level. Scoring criteria for level 4 and 5 are more stringent than criteria for an A in most classes, and scores of 4 and 5 must be defended during the scoring process to a greater degree than scores of 3.

It is highly likely that the correlation between proficiency scores and grades represents the degree to which the two systems assess certain core academic skills, such as writing and mathematical reasoning. Another possibility is that the correlation represents the "G-factor"—that portion of test scores explainable based on generalized intelligence. The latter hypothesis is less likely given the relative homogeneity of the test taking population, whereas the former is more plausible.

One other observation is worth considering. The data make a case for the existence of grade inflation as a real phenomenon. The scoring process that was created for the collections of student work was consciously pegged to the skills needed to do entry-level university work successfully. The standards and criteria were linked directly with this outcome. Based on this standard, only the students who were being awarded A's in high school were highly likely to meet the standard, and even within this group, sizeable numbers of students who received A's did not receive scores of 3, the minimum level for admission.

A number of possible explanations can be offered, including the newness of the proficiency system and its lack of a direct effect on students. However, the pieces of work submitted to be judged were also graded by teachers and contributed to class grade, so students likely took the work seriously. That they were able to earn A's and not be judged proficient is cause for further investigation. That students receiving B's had an average score well below the proficient level perhaps bears even more investigation.

Similarly, the lack of any systematic relationship between the grading system teachers used and the proficiency score students received suggests a disconnect between the proficiency-based scoring system, which focuses only on student performance, and the grades, which apparently capture more varied aspects of the classroom experience. One might reasonably expect that grading systems that emphasized, say, written assignments in English might have been more closely associated with students judged proficient in English, but this was not the case. Granted, teacher self-reports of grading systems are approximate measures. But the work collections that were scored contained pieces that were also graded. It is reasonable to expect more of a connection between the two. The implication is that even when teachers focus their grading in areas that produce work that is similar to what is required for the proficiency collections (e.g., papers, tests, projects), teachers apparently do not grade this work in a way that is consistent with how it would be judged externally by those trained to apply proficiency standards related to university admission.

It could well be that the proficiency standards are too high. However, these standards have been developed over a six-year period with constant input and review by hundreds of high school teachers and university faculty. These standards should be as close to an accurate statement of the mutual expectations high school teachers and university faculty have for college-bound students as is possible to achieve currently. If this is the case, the gap between the proficiency scores and the grades students are receiving suggests that grade inflation is real, significant, and not adequately recognized as a wide-spread phenomenon in American high schools.

Plans are to repeat this research with students whose work will be judged in May, 2000 and to gather information on the scores these students receive on their PSAT and SAT tests as well as on state tests in mathematics, math problem solving, writing and English. This broader set of measures will help establish the concurrent validity among these various ways of judging student college-readiness. The contribution that proficiency scores can make to university admissions decisions (as well as the limitations of existing methods) can be better considered when the relationships among the measures in better understood. This study was a first attempt to explore those relationships and to consider grades within a different, external context. The results presented here suggest that further study is justified and necessary to determine how best to utilize teacher judgment of student work— through individualized grading systems or via common standards applied in a consistent fashion to all students.

References

Gipps, Caroline V. (1994). Beyond Testing: Towards a Theory of Educational Assessment. London: Falmer Press.

Gose, Ben (1997, July 25). Efforts to Curb Grade Inflation Get an F From Many Critics. Chronicle of Higher Education. 43(46), A41-A42.

Guskey, Thomas (1994, October). Making the Grade: What Benefits Students? Educational Leadership. 52(2), 14-20.

Marzano, Robert J. (1994). Lessons from the field about outcome-based performance assessments. Educational Leadership. 75(6), 44-50.

Minneapolis Star Tribune (1998, August 4). Schools seek new gauges for grading. Minneapolis Star Tribune, p. 1A.

Ornstein, A. C. (1994). Grading Practices and Policies: An Overview and Some Suggestions. NASSP Bulletin. 78(559), 55-64.

Sadler, D. Royce (1992, November). Expert Review and Educational Reform: The Case of Student Assessment in Queensland Secondary Schools. Australian Journal of Education. 36, 301-318.

Sadler, R. (1987). Specifying and promulgating achievement standards. Oxford Review of Education. 13(2).

Seeley, Marcia M. (1994, October). The Mismatch Between Assessment and Grading. Educational Leadership. 52(2), 4-6.

Ziomek, Robert L. & Joseph C. Svec (1995, November). High School Grades and Achievement: Evidence of Grade Inflation. American College Testing Program, Iowa City, Iowa.